City of San Antonio
Open Data Procedures

Updated on: August 1, 2018

Purpose of this document:

To issue guidelines for internal departments to identify, review, prioritize and prepare publishable City of San Antonio open data for access by the public via the OpenGov Open Data Portal. The procedures listed in this document are applied in conjunction with the Open Data Policy of the City of San Antonio and its Administrative Directives 7.3A

Open Data Portal

The Open Data Portal is hosted by OpenGov, administered by the IT Department, and shall act as a central data hub for all open data published by and on behalf of the City of San Antonio.

The internet site created for this purpose is currently located at:

https://data.sanantonio.gov/.

The concept of "Open Data" describes data that is freely available, machine-readable, and formatted according to national technical standards to facilitate visibility and re-use of published data.

The OpenGov Open Data Portal was developed to catalog data and enable data to be readily discoverable. The portal offers access to standardized data that can be easily retrieved, downloaded, sorted, searched, analyzed, redistributed and re-used by individuals, business, researchers, journalists, developers, and government to process, trend, and innovate utilizing a single data Resource or combinations of data Resources.

Data posted to the Open Data Portal shall meet the following criteria:

- Readable, indexable, electronically searchable and discoverable by commonly used Internet search applications
- Platform independent and machine readable
- Available to the public free of charge

Selecting data to be published in the Open Data platform:

A **Data Coordinator** shall lead the data preparatory process. This individual will facilitate initial solicitation for data resources, suggestions for publication and subsequent preparation activities among the department's centers, divisions and offices.

In addition, any data steward in a department can and should consider identifying data that could fulfill strategic needs by sharing them on the Open Data Portal.

After identification, all suggested data resources should be assessed and prioritized.

Each department willing to publish open data should establish an internal review and selection process to obtain approval for data resources to be published.

Departments and offices are responsible for driving toward increasing data content, quality, and accuracy, as well as ensuring compliance with all security, privacy, confidentiality laws, rules, regulations, and intellectual property rights requirements. (Administrative Directive 7.3A refers.)

Data Resources Assessment and Prioritization

In prioritizing data for release, departments and offices must account for time to: identify data, assess and validate the data (i.e., ensure consistency, timeliness, relevance, completeness, and accuracy of the data), ensure completeness of the metadata and data dictionary, prepare visualizations and talking points, and obtain all necessary approvals to publish the data.

High value data is that one that can be used to increase the City's accountability and responsiveness, improve public knowledge of the City and its operations, further its mission, create economic opportunity, or respond to a need or demand identified after public consultation.

There may be statutorily required reporting which can be satisfied by publishing data resources, without necessarily producing an additional extensive narrative report. If the data is collected and compiled by the department to fulfill statutory reporting requirements, then the department's governing laws have already determined that the data is of high value for that department.

Data resources will be formatted in a machine-readable format. The City has chosen Comma Separated Values (CSV) as its standard format for publication. Accompanying the data resources will be complete metadata and a data dictionary that provides descriptions and technical notes as necessary for every field in the data resource.

Departments and Offices are also encouraged to include visualizations of the data (graphs and/or maps) as a way to encourage public engagement and innovation related to strategic goals.

Publication of Open Data

Each data resource being published on the portal requires appropriate categorization and tags (key words) to provide ease in searching for the data content.

Furthermore, Departments and Offices may consider for each dataset sharing its publication via social media, a press release or other communication method.

The department or office Data Coordinator is responsible for obtaining the following approvals from within their department:

1. Data Steward: Refer to description in the Open Data SA policy document.
2. Director: The head of the department or office (or their designee, such as the Data Coordinator) validates that the center/division/office wishes to proceed with publication of the data resource and assumes responsibility for the global review of the data including an evaluation of sensitivities that may be associated with it.
3. Legal Counsel: Legal counsel will be in the best position to determine, when needed, whether the data resource has internally been reviewed sufficiently to ensure compliance with privacy and security requirements, intellectual property rights, and Public Records Act (PRA) responsibilities. Legal counsel may recommend additional consultation with the chief security officer, and/or public affairs officer.

Once selected data and approvals have been obtained, contact the Data analytics Team in the IT Department to express your interest in using the Open Data Platform, identify the data source and establish a data upload schedule.

Metadata

The City's Open Data adheres to core components of the Project Open Data schema for metadata (https://project-open-data.cio.gov/v1.1/schema/).

The ability to search and find information is enhanced by the adherence to metadata standards required with each dataset. Metadata includes subject categories and keywords which provide for more precise searching and document management. Adoption of the Project Open Data Metadata schema, maximizes adaptability and interoperability. Metadata fields are listed in the Open Data Policy.

Open Data-OpenGov Portal Roles

In order to be able to have access to the Open Data Portal to publish, edit or receive data notifications, a role needs to be assigned to the data user.

Available roles in the platform are as follows:

Admin	Can Do Everything As Editor Plus: Add Users To The Organization, Assign The User Role, Remove Members, Editors Or Other Admins; Edit Or Delete Organizations. This role has been reserved for the Open Data Architect and the Data Analytics Team in the IT Department.
Editor	Can Do Everything As Member Plus: Add New Datasets, Edit, Delete Or Make Datasets Public Or Private. This role can be assigned to the Data Stewards or Coordinators of the participating Departments.
Member	Can View The Organization's Private Datasets.
Registered User	Can get an API key and follow datasets to receive updates

Data Resources

The COSA Open Data Portal supports two classifications of data resources: tabular and geospatial. A tabular data resource is a flat file that conforms to a predefined schema. The schema defines the characteristics of a fixed number of columns, including the column name and data type. A geospatial data resource contains information that can be readily rendered on an underlying map. Examples of geospatial features include points (buildings), polylines (bus routes), and polygons (school districts), along with attribute information that describes characteristics of each spatial feature.

Tabular
Data resources can be exported for download in popular human-readable formats, machine-readable standards and streamable file formats. The COSA Open Data Portal currently supports the following exportable tabular file formats:

- CSV
- JSON
- PDF
- RDF
- RSS
- XLS
- XLSX
- XML

Geospatial
Geospatial data contain geographic feature and attribute data that define the properties of geographic features which may be used in a geographic information system (GIS). Attributes are stored in a tabular format with unique key references to the associated geographic features. Two export methodologies are supported for geographic information: geospatial and attribute. Attribute layers can be exported as tabular data file formats (see tabular formats listed above). Geospatial data can be downloaded in any of the tabular formats defined above, as well as the following formats:

- .Shapefile
- .Keyhole Markup Language (KML/KMZ)

Large Files
Public data often consist of historical archives, comprised of potentially millions of records collected over an extended period of time. The CHHS Open Data Portal supports the loading, exporting and visualization of large data resources (> 1GB).

Security, Privacy, Regulatory & Aggregate Data

The public release of some department data might result in the violation of laws, rules, or regulations. Some data may not be appropriate to release because it can compromise internal departmental processes, such as procurement. Other data may contain **personally identifiable information**. Finally, even if detailed data appear innocuous, it may be possible to combine it with other public information to reveal sensitive details (commonly known as the mosaic effect). Before disclosing potential personally identifiable information or other potentially sensitive information, departments and offices must make a

'best effort' to consider other publicly available data – in any medium and from any source – to determine whether some combination of existing data and the data intended to be publicly released present any risks or would make the publication inappropriate. (Administrative Directive 7.3A refers).

Removing Published Data from platform:

If a data resource needs to be removed from the platform, the Open Data Architect and the Data Analytics Team in the IT departments need to be contacted immediately to ensure prompt data removal from public view.

You can call the Help Desk at 7-888 and they will transfer your request to the appropriate Open Data Support Group. You can also send email request to _____@_____.

Glossary

API
An application programming interface, which is a set of definitions of the ways one piece of computer software communicates with another. It is a method of achieving abstraction, usually (but not necessarily) between higher-level and lower-level software.

Catalog
A catalog is a collection of data resources or web services.

CSV
A comma separated values (CSV) file is a computer data file used for implementing the organizational tool, the Comma Separated List. The CSV file is used for the digital storage of data structured in a table of lists form. Each line in the CSV file corresponds to a row in the table. Within a line, fields are separated by commas, and each field belongs to one table column. CSV files are often used for moving tabular data between two different computer programs (like moving between a database program and a spreadsheet program).

Data
A value or set of values representing a specific concept or concepts. Data include, but are not limited to, 1) geospatial data 2) unstructured data, and 3) structured data.

Data is Publishable City Data if it meets one of the following criteria:
(1) Data that is public by law such as via the Public Records Act or
(2) Data that is not prohibited from being released by any laws, regulations, policies, rules, rights, court order, or any other restriction.

Data shall not be released if it is highly restricted due to the Health Insurance Portability and Accountability Act ("HIPAA"), state, local or federal law.

Database
A collection of data stored according to a schema such that a computer can easily find the desired information.

Dataset
An organized collection of related data records maintained on a storage device, with the collection containing data organized or formatted in a specific or prescribed way, often in tabular form. A dataset refers to the master, primary, or original authoritative collection of the data

Data Resource
A data resource, refers to a subset of the dataset which may include a selection and/or aggregation of data from the original dataset.

Geographic Information System (GIS)
A geographic information system (GIS) is a computer system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data allowing the user to question, analyze, and interpret data to understand relationships, patterns, and trends.

Information
Information means any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms.

JSON
JSON (JavaScript Object Notation) is a lightweight data-interchange format. It is easy for both humans to read and write and machines to parse and generate. It is based on a subset of the JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999. JSON is a text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and many others. These properties make JSON an ideal data-interchange language.

Machine-Readable File
Refers to information or data that is in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost.

Metadata
Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource (NISO 2004, ISBN: 1-880124-62-9). The more information that can be conveyed in a standardized regular format, the more valuable data becomes. Making metadata machine readable greatly increases its utility, but requires more detailed standardization, defining not only field names, but also how information is encoded in the metadata fields.

Public Information Act
The process by which the public requests state or local government records.

Publishable City Data
A data resource that meets one of the following criteria: (1) data that are public by law such as via the Public Records Act or (2) the data is not prohibited from being released by any laws, regulations, policies,

rules, rights, court order, or any other restriction. Data shall not be released if it is highly restricted due to HIPAA, state, or federal law.

Unstructured Data
Data that is more free-form, such as multimedia files, images, sound files, or unstructured text. Unstructured data does not necessarily follow any format or hierarchical sequence, nor does it follow any relational rules. Unstructured data refers to masses of (usually) computerized information which do not have a data structure which is easily readable by a machine. Examples of unstructured data may include audio, video and unstructured text such as the body of an email or word processor document. Data mining techniques are used to find patterns in, or otherwise interpret, this information.

Catalogs

Datacatalogs.org | http://datacatalogs.org/
DataCatalogs.org aims to be the most comprehensive list of open data catalogs in the world. It is curated by a group of leading open data experts from around the world - including representatives from local, regional and national governments, international organizations such as the World Bank, and numerous NGOs.

LOGD Dataset Catalog | http://logd.tw.rpi.edu/datasets
The Linking Open Government Data (LOGD) project investigates opening and linking government data using Semantic web technologies. We are translating government-related datasets into RDF, linking them to the Web of Data and providing demos and tutorials on mashing up and consuming linked government data.

Open Data Portals

By City

Birmingham, AL: https://data.birminghamal.gov/about

Boston, Massachusetts https://data.cityofboston.gov/

CHHS: https://data.chhs.ca.gov/pages/terms

Chicago, Illinois https://data.cityofchicago.org/

Dallas, Texas https://www.dallasopendata.com/

Lexington, Kentucky http://data.lexingtonky.gov/

Louisville, Kentucky http://portal.louisvilleky.gov/service/data

Madison, Wisconsin https://data.cityofmadison.com/

New York, New York https://nycopendata.socrata.com/

Oakland, California https://data.oaklandnet.com/

Philadelphia, Pennsylvania http://www.opendataphilly.org/

Phoenix, AZ: https://phoenixopendata.com/pages/terms-of-use

Raleigh, North Carolina https://data.raleighnc.gov/

Sacramento, California http://data.cityofsacramento.org/home/

San Francisco, California https://data.sfgov.org/

Seattle, Washington https://data.seattle.gov/

South Bend, Indiana https://data.southbendin.gov/

Tempe, AZ: https://data.tempe.gov/about

By County

Alameda County, California-https://data.acgov.org/

Cook County, Illinois, https://datacatalog.cookcountyil.gov/

San Mateo County, California, https://data.smcgov.org/

Strathcona County, Alberta, Canada, https://data.strathcona.ca/

Wake County, North Carolina, http://www.wakegov.com/data/Pages/default.aspx

By State

Connecticut, USA, https://data.ct.gov/

Hawaii, USA, https://data.hawaii.gov/

Illinois, USA, https://data.illinois.gov/

Maryland, USA, https://data.maryland.gov/

Oregon, USA, https://data.oregon.gov/

By Country

Canada, http://data.gc.ca/eng

United States, https://www.data.gov/

European Union, http://open-data.europa.eu/en/data/

India, http://data.gov.in/

Kenya, https://opendata.go.ke/

UK Government, http://data.gov.uk/

United Nations, http://data.un.org/

World Bank, http://data.worldbank.org/

Reference Materials

Amazon Web Services public datasets http://aws.amazon.com/datasets Huge resource of public data, including the 1000 Genome Project, an attempt to build the most comprehensive database of human genetic information and NASA 's database of satellite imagery of Earth.

Buzzdata is a social data sharing service that allows you to upload your own data and connect with others who are uploading their data.

The **CIA** World Facebook https://www.cia.gov/library/publications/the-world-factbook/ Information on history, population, economy, government, infrastructure and military of 267 countries.

Data Market is a place to check out data related to economics, healthcare, food and agriculture, and the automotive industry.

DBPedia http://wiki.dbpedia.org Wikipedia is comprised of millions of pieces of data, structured and unstructured on every subject under the sun. DBPedia is an ambitious project to catalogue and create a public, freely distributable database allowing anyone to analyze this data.

Facebook FB -0.78%

Face.com: A fascinating tool for facial recognition data.

Financial Data Finder at OSU offers a large catalog of financial data sets.

Freebase http://www.freebase.com/ a community-compiled database of structured data about people, places and things, with over 45 million entries.

Gapminder http://www.gapminder.org/data/ Compilation of data from sources including the World Health Organization and World Bank covering economic, medical and social statistics from around the world.

Github | http://project-open-data.github.io/
Powerful collaboration, code review, and code management for open source and private projects.

Google GOOGL -2.55%

Google Finance https://www.google.com/finance 40 years' worth of stock market data, updated in real time.

Google Books Ngrams http://storage.googleapis.com/books/ngrams/books/datasetsv2.html Search and analyze the full text of any of the millions of books digitized as part of the Google Books project.

Google Public data explorer includes data from world development indicators, OECD, and human development indicators, mostly related to economics data and the world.

Healthdata.gov https://www.healthdata.gov/ 125 years of US healthcare data including claim-level Medicare data, epidemiology and population statistics.

Million Song Data Set http://aws.amazon.com/datasets/6468931156960467 Metadata on over a million songs and pieces of music. Part of Amazon Web Services.

National Climatic Data Center http://www.ncdc.noaa.gov/data-access/quick-links#loc-clim huge collection of environmental, meteorological and climate data sets from the US National Climatic Data Center. The world's largest archive of weather data.

NHS Health and Social Care Information Centre http://www.hscic.gov.uk/home Health data sets from the UK National Health Service.

New York Times NYT -1.84%

Open Data Handbook, http://opendatahandbook.org/resources/

Pew Research Center offers its raw data from its fascinating research into American life.

Sunlight foundation: https://sunlightfoundation.com/.The Sunlight Foundation is a national, nonpartisan, nonprofit organization that uses civic technologies, open data, policy analysis and journalism to make our government and politics more accountable and transparent to all.

The BROAD Institute offers a number of cancer-related datasets.

UCI Machine Learning Repository is a dataset specifically pre-processed for machine learning.

UCLA makes some of the data from its courses public.

UNICEF offers statistics on the situation of women and children worldwide.

US Census Bureau http://www.census.gov/data.html a wealth of information on the lives of US citizens covering population data, geographic data and education.

World Health Organization offers world hunger, health, and disease statistics.

Solutions

CKAN | http://ckan.org/solutions/
CKAN is the world's leading open-source data portal platform. It is a complete out-of-the-box software solution that makes data accessible – by providing tools to streamline publishing, sharing, finding and using data.

ESRI | https://opendata.arcgis.com/about
ArcGIS Open Data allows you to leverage your investment in the ArcGIS platform. You'll be able to easily share the data you've collected, curated, and maintained in ArcGIS Online with everyone.

Junar | http://www.junar.com/
Junar delivers all the benefits of SaaS (Software-as-a-Service) to help organizations Open Data to spur innovation. Junar makes is easy to deal with complex end-to-end Open Data projects and turns the difficult task of opening data into a secure and controlled process.

OpenGov | https://opengov.com/open-data
OpenGov's comprehensive open data and financial transparency solutions help agencies of all sizes drive accountability, make data more useful, engage the public, and unlock economic potential.

Socrata | http://www.socrata.com/
Socrata helps public sector organizations improve transparency, citizen service, and data-driven decision-making. Socrata's user-friendly solutions deliver data to governments trying to reduce costs, to citizens who want to understand how their tax dollars are used, and to civic hackers dedicated to creating new apps and improving services.

Visualizations

EPA Data Finder | http://www.epa.gov/emefdata/em4ef.home
This site makes available a large selection of EPA data sources, organized into topics such as air and water that are in easily downloadable formats. Data Finder points to data in downloadable formats to speed up environmental research. For each data source, you can see a basic overview, including the geographic scale and other contextual information, then access the data source itself.

Geographic Information System at CDC | http://www.cdc.gov/gis/index.htm
GIS at CDC was developed to create and display interactive maps of national, state, or country rates for key health statistics originating from different CDC programs, broken out by geography, gender, and ethnicity. Web applications include, but are not limited to, Environmental Health, Chronic Disease Prevention and Health Promotion, and Rabies Surveillance.

Health Map | http://www.healthmap.org/en/
HealthMap brings together disparate data sources to achieve a unified and comprehensive view of the current global state of infectious diseases and their effect on human and animal health. This freely available site integrates outbreak data of varying reliability, ranging from news sources to curated personal accounts to validated official alerts.

National Cancer Institute GIS & Science | http://gis.cancer.gov/
This site serves as a central source of information, data, tools, relevant publications and web mapping tools for cancer data.

National Environmental Public Health Tracking Network |
http://ephtracking.cdc.gov/showHome.action
The National Environmental Public Health Tracking Network (Tracking Network) is a system of integrated health, exposure, and hazard information and data from a variety of national, state, and city sources. On the Tracking Network, you can view maps, tables, and charts with data about Environments, Public Health, Health Effects, and info by locations.

Policy Map | http://Policymap.com
PolicyMap is an online data and mapping tool that enables government, commercial, non-profit and academic institutions to access data about communities and markets across the US. Use it for research, market studies, business planning, and site selection, grant applications and impact analysis.

TOXMap – Environmental Health E-Maps | http://toxmap.nlm.nih.gov/toxmap/main/index.jsp
TOXMAP is a GIS from the Division of Specialized Information Services of the US National Library of Medicine that uses maps of the United States to help users visually explore data from the US Environmental Protection Agency (EPA)'s Toxics Release Inventory (TRI) and Superfund Program.

World Health Organization Data & Statistics | http://gis.emro.who.int/PublicHealthMappingGIS/
The World Health Organization (WHO) maintains mortality and global health estimates and provides access to data for analysis and monitoring through a data depository, world health statistics report, country statistics, maps, and the WHO indicator registry. WHO has also created a HealthMapper application to address critical surveillance needs across infectious disease programs at national and global levels.